DOUGLAS

MISSILE & SPACE SYSTEMS DIVISION / ASTROPOWER LABORATORY
2121 PAULARINO AVENUE, NEWPORT BEACH, CALIFORNIA

Technical Paper No. 1967

CLOUD PATTERN RECOGNITION

February 1964

Prepared By

R. D. Joseph
S. S. Viglione
H. F. Wolf

DOUGLAS AIRCRAFT COMPANY, INC. SANTA MONICA, CALIFORNIA

CLOUD PATTERN RECOGNITION

R. D. Joseph, S. S. Viglione, H. F. Wolf*

ABSTRACT

Cloud cover photographs transmitted from meteorological satellites must be processed and interpreted before weather maps can be issued. Most of the routine processing can be handled by present day digital computer techniques; however, the recognition and interpretation of cloud patterns, such as vortices indicating hurricanes, must still be performed by humans due to the lack of suitable recognition mechanisms. This paper investigates the feasibility of using a perceptron-type computer for the recognition of vortex patterns. A formula is derived which enables the prediction of machine performance as a function of problem complexity and perceptron size (number of logic units). It is shown that the problem complexity can be estimated through optical correlation measurements on cloud cover negatives. These measurements are described and a computer routine is developed which mechanizes the prediction equations and examines the experimental data gained from 10,000 measurements. The results of the computer program are presented and their meaning is discussed.

*Astropower Laboratory, Missile & Space Division, Douglas Aircraft Co., Inc.

INTRODUCTION

This paper discusses the results of a study* to determine the feasibility of using a self-organizing, parallel logic system, as diagrammed in Figure 1, for the recognition of cloud patterns.

The input connections for each majority logic unit sample the cloud pattern in storage. The logic units compute majority logic propositions concerning the cloud pattern, the particular proposition being determined by the input connections. Thus each majority logic unit becomes a property filter, recognizing a specific property of the incoming pattern. The input connections themselves represent a sampling of a much larger list of property filters. This list may be selected completely at random, determined entirely from the known properties of the patterns, or compiled by combinations of these approaches - the method of selecting the property filter list depends upon how well the bases for pattern classification are understood. The self-organizing routine determines the utility of the various property filters included in the machine.

Feasibility of this approach was to be determined by estimating the number of logic units required for the recognition of vortices. Theoretical optimization studies were also undertaken to reduce the complexity of the logic layer by increasing the efficiency of each logic unit and by optimizing the logic input connections and threshold setting.

A review of self-organized pattern recognition machines is presented. The general machine organization, the learning rules applied during the self-organizing period, and several procedures for reducing the size of machine required for a given performance are discussed. The next section gives the mathematical derivations upon which the estimation of machine size is based. This is followed by a discussion in detail of the selection and preparation of the cloud photographs, the design of the optical correlator, and the optical measurements. The final section discusses the computer program that was written to evaluate the optical data and gives the results of the program.

---

*This work was performed in part under contract NASw609 from the Directorate of Electronics and Data Processing, NASA Headquarters, Washington, D. C.

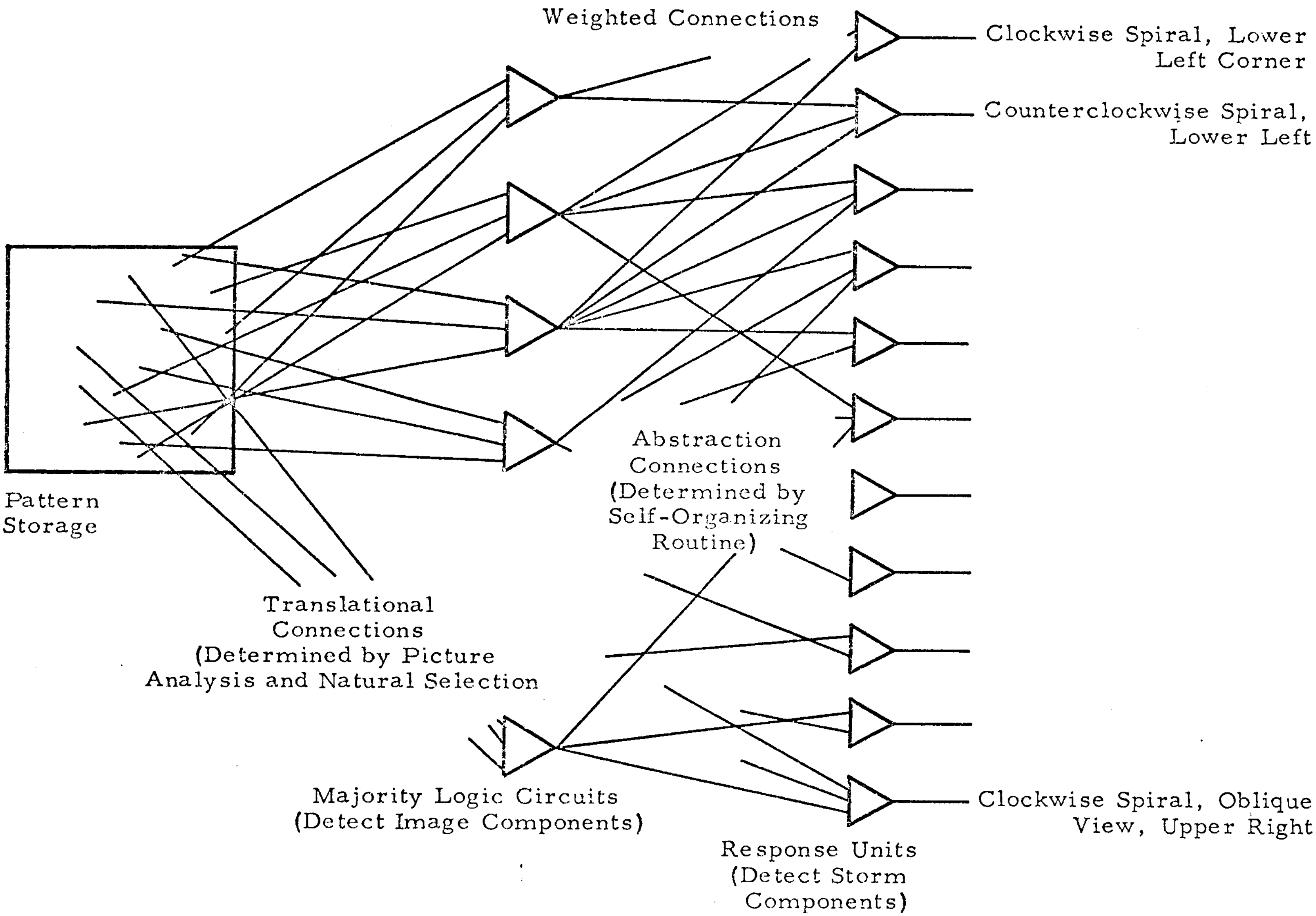


Figure 1. Block Diagram of Parallel Logic Interpreter

SELF-ORGANIZING, PARALLEL LOGIC SYSTEMS

Pattern classification must often be based on incompletely specified criteria empirically derived from "typical examples" of the pattern classes. While this method is characteristic of humans, it has only recently been adapted to automata. Machines which use this method are called self-organizing systems.

Organization of Parallel Logic Systems

In the devices to be considered, the patterns are to be classified on the basis of a discrete sampling of the pattern data. The samples obtained from a particular pattern class may be conceptualized as a set of coordinates which represents the pattern as a point in a k-dimensional signal space. The design of an automatic pattern classification device then is largely the specification of a partition on this space such that the cells of the partition may be identified with unique classes of patterns. The design constraints must include the maximum probability of error acceptable after the specification of the partition, and the limitations on the cost necessary to automate the process of specifying the classification regions (the decision process).

When the patterns associated with unique classes are concentrated in well-defined and widely separated regions of the signal space, simple and highly accurate decision mechanisms may be constructed. In more difficult designs, a useful approach is to employ simple decision devices (to achieve low cost) in fairly elaborate structures (to achieve the required partition complexity). Such a structure is diagrammed in Figure 2.

The "S" units are the sensory or input units to the device. In the case of pictorial data, a sensory unit is associated with a picture element and generates a signal proportional to the brightness of that element. The sensory field is connected to the first-layer logic units through weighted connections to translate the incoming signal to a form more suitable for recognition. The input signal may be translated through several layers of logic units, but eventually there is an inevitable "necking down" of the data - referred to as abstraction.

The performance of a parallel logic system may be characterized in terms of the response units. The response units considered here are two-state

TP1967

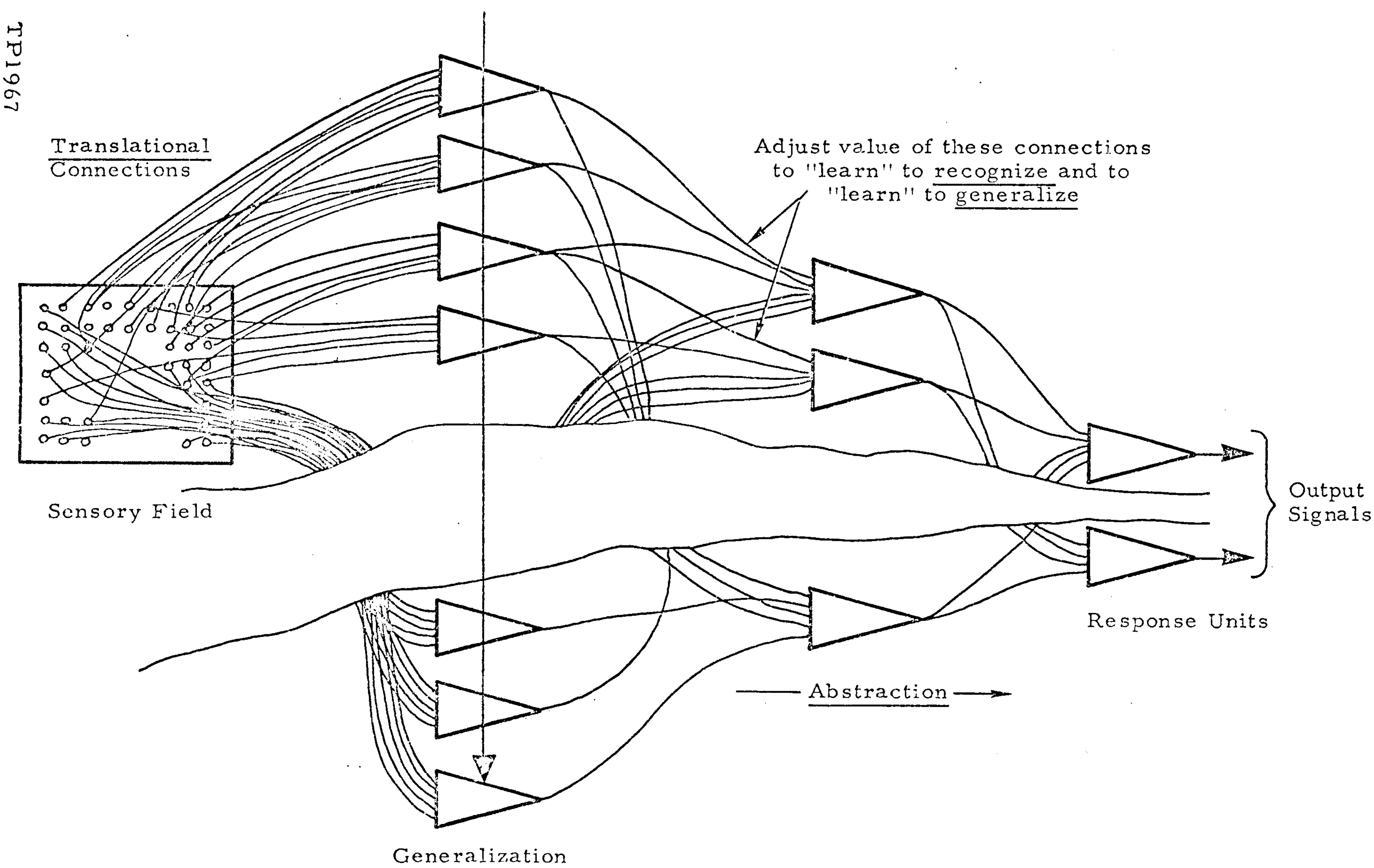


Figure 2. Generalized Parallel Logic, Distributed Memory System

devices required to be in one state when a particular pattern is present in the signal displayed to the sensor units, and in a second state when this pattern is not present.

Initially there is little organization in these machines. However, by showing a sequence of signals to the machine and varying the weights of the connections between logic layers, or between logic and response layers, the machine's ability to classify patterns is improved.

A significant feature of these machines is their generalization capability. Since the classification of a pattern is made on a statistical basis, many logic units are active for each correct classification. The existence of a slightly different property in a signal of the same class as that upon which the machine was trained activates a large portion of the appropriate logic units, enabling correct classification despite minor variations in the pattern.

The decision units in Figure 2 are linear logic units.* A linear logic unit partitions the subspace defined by the sources of its inputs by passing a single hyperplane through the subspace, assigning a classification of "one" to patterns falling on one side of the hyperplane and "zero" to patterns falling on the other side. The linear logic unit mechanism is diagrammed in Figure 3. The rule governing the behavior of this unit is as follows: if the weighted sum of the input $\sum_i w_i e_i$ exceeds a threshold, θ, then an output, e_o, is generated and the unit is said to be active.

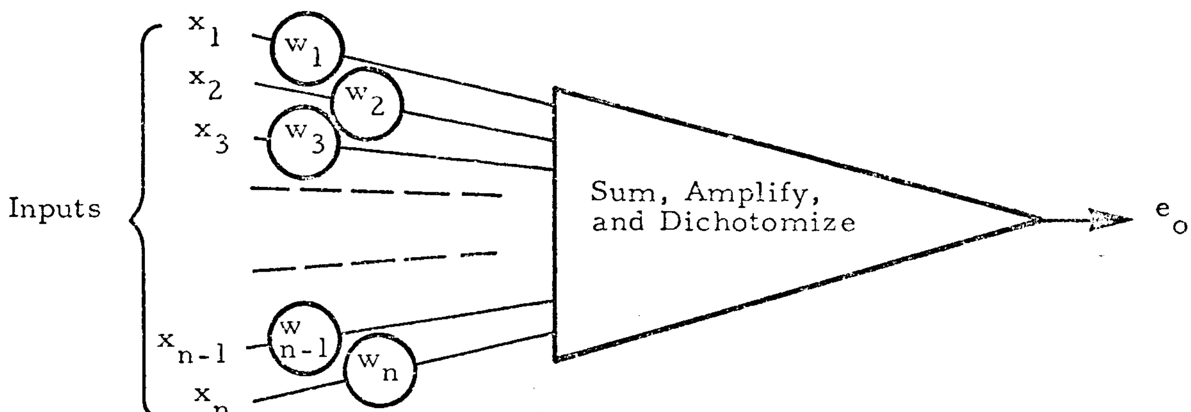
$$e_o = \begin{cases} 1 \text{ if } \sum_i w_i e_i \geq \theta \\ 0 \text{ otherwise} \end{cases}$$

The simplest organization which would give the machine a capacity to perform significant tasks contains two layers of logic (Figure 4).** The sensory field again provides the input, and a linear logic unit serves as a report-out unit. Between the two is inserted a layer of linear logic units. All of the connections

---

* Winder, R. O., "Threshold Logic in Artificial Intelligence," Artificial Intelligence Sessions, IEEE Winter General Meeting, Jan 1963.

** It should be noted that many pattern recognition devices are based on this structure - for example, the designs of Bledsoe and Browning, Gambs, Kamensky, Rosenblatt, Widrow, and the Astropower Decision Filter.



$$e_o = \begin{cases} +1 \text{ if } \sum_{j=1}^{n} w_j x_j > 0 \\ 0 \text{ otherwise} \end{cases}$$

Figure 3. Linear Logic Unit Mechanism

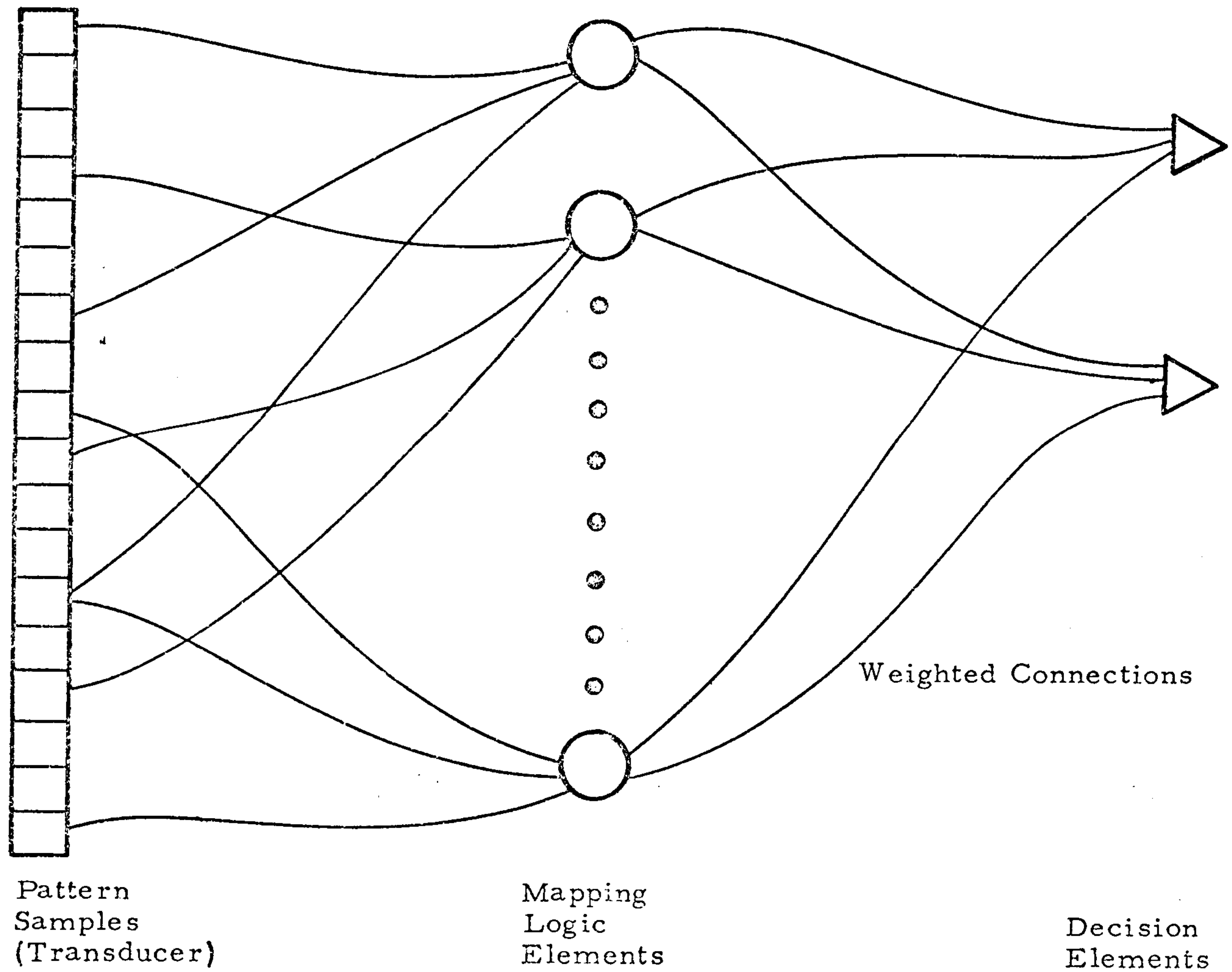


Figure 4. Pattern Recognition Network Organization

shown are weighted. The connections from sensory points to logic units are many-to-many, and from logic units to report-out units are all-to-all.

Self-Organizing Algorithms

Forced Learning

After selecting this organization as the network structure, a large number of network parameters remain to be specified. These parameters include the number of input connections, and the source points of these inputs for each logic unit; the weights of all connections from the sensory field to the logic units, and from the logic units to the decision elements; and the thresholds of all logic units and decision elements. A rule for automatically determining the values of these parameters based on a sample of classified patterns is called a self-organizing algorithm.

The particular algorithm utilized is called "forced learning" and is implemented in the following manner. When a logic unit is supplied with a set of input connections, it becomes a property filter,* determining whether or not the input patterns possess the particular property. A list of properties is established, and a sampling of these is selected for the recognition machine. The self-organizing procedure then determines the utility of the various property filters and the efficiency of the property list in making the required classification. The learning rule is such that it increases the weights of the connections from active units to the report-out unit when a pattern of interest is present, and reduces the connection weight of active units when no target or pattern of interest is present. With such a procedure, reliable indicators of target presence (that is, units which measure properties possessed by patterns of interest) will tend to carry large positive weights. Units which are reliable indicators that a target is not present, and hence measure properties possessed primarily by nontargets, will tend to carry large negative weights. Units which are less reliable indicators of the presence or absence of patterns of interest will carry positive or negative weights of lesser magnitude.

---

*Singer, T. R., "Model for a Size Invariant Pattern Recognition System," paper presented at the Bionics Symposium, Dayton, O., Sept 1960; Martin, T. B., and Talavage, J. F., "Application of Neural Logic to Speech Analysis and Recognition," paper presented at the Bionics Symposium, Dayton, O., 1963.

The list of property filters from which the logic unit input connections are selected is quite significant. The structure of this list determines the task for which the machine is suitable - that is, if the recognition machine is to perform properly, it must have the foundations upon which to function. Many techniques are available for generating this property list. One method is to manually build into the list each known parameter of the patterns to be classified. Such a procedure is time-consuming and tedious, and above all, it may not lead to a machine capable of correct classification if the discrimination clues are subtle and not known to the designer. One approach to this problem is to incorporate those properties known to be useful, then generate additional properties randomly and let the self-organizing procedure select the discrimination criteria. Such a procedure requires an extensive property list and may be accommodated by performing the self-organization in a simulated mode on a large scale general purpose digital computer.

The forced learning procedure is readily implemented in a computer simulation. It is well documented mathematically and gives rise to the "Theorem of Statical Separability."* For this theorem to hold, it is required that the property list emphasize the difference between pattern of interest and pattern of no interest, and the similarities between these classes. The theorem makes it possible to compute whether or not the property list meets this requirement and to establish the required size of the logic layer for a specified performance level. It formed the basis for the work performed and is discussed thoroughly in a following paragraph entitled "Estimation of Machine Size."

<u>Alternate Algorithms</u>

Machine designs achieved with forced learning usually require excessive numbers of logic units since all those that were originally generated, no matter what their contribution to the classification task, are included in the final design. Other algorithms have been developed for the design of efficient networks by increasing the effectiveness and the efficiency of the logic structure. One such

*Joseph, R. D., "Contributions to Perceptron Theory," Ph. D Thesis, Cornell University, Sept 1961.

procedure involves the examination of the assigned weights, eliminating those units that have weights of zero or negligible magnitude and retain only those units with the largest positive and negative value. This is known as the "natural selection"* principle and was developed specifically to decrease the complexity of machines designed wiith forced learning.

A particularly powerful algorithm,** developed concurrently with this program, employs the statistical techniques of discriminant analysis to generate populations of logic units to be examined, while using a minimum loss criteria for determining the suitability of these units for inclusion in the final network. This "iterative design" routine shows evidence of being able to achieve extremely efficient designs and of effectively separating complex pattern classes.

---

* Joseph, R. D., Kelly, P. M., Viglione, S. S., "An Optical Decision Filter," Proc. IEEE, Vol. 51, August 1963.

** Daly, J. A., Joseph, R. D., Ramsey, D. M., "An Iterative Design Technique for Pattern Classification Logic," presented at WESCON, August 1963.

ESTIMATION OF MACHINE SIZE

Statistical analysis of perceptron-type systems permits the estimation of their performance capabilities without actual network construction or simulation. When the recognition task becomes even moderately complex, however, the exact statistical analyses become unwieldy. Astropower personnel have developed an approximate analysis which is in fact less cumbersome then simulation. This section presents an exact analysis, the approximation, and means for estimating the parameters required.

Exact Perceptron Analysis[1,2]

The problem to be considered is the classification of a large number of multidimensional vectors into two categories. The data in each signal point are to be processed by a large number of logic units simultaneously; each logic unit determines the truth of some linear logic proposition when applied to the vector under examination. As a result of this detailed examination of the individual vector, that vector receives a classification. Suppose that the sensory field under consideration contains n points, so that the weighted input connections to a given logic unit can be represented as an n-dimensional vector. The j-th component of this vector represents the weight of the input connection from the j-th sensory point. The absence of a connection is represented by a zero.

The connection vectors for the logic units are selected randomly as follows. A distribution is assigned to the space of all possible input connection vectors (n-space), and the connection vector for each logic unit is selected independently according to this distribution. The set of connection vectors assigned non-zero probability is called the property list. In this case, the property list was generated by assuming that the sum of the weights of the input connections to each logic unit is "A," and the sum of squares of these weights is "B" (the numbers A and B being the parameters to be optimized), and that the points of origin of each input connection are selected independently, according to a uniform distribution over the entire sensory field.

[1] Joseph, R. D., "Contributions to Perceptron Theory," Cornell Aeronautical Laboratory Report VG 1196-G7.

[2] Rosenblatt, F., _Principles of Neurodynamics_, Spartan Books, 1962

Let C_i denote the input connection vector for the i-th logic unit, and S^k the input pattern expressed as an n-dimensional vector. The output of the i-th logic unit when the k-th pattern is shown, which will be denoted by δ_i^k, can be expressed in terms of the inner product of the connection vector and the signal vector

$$\delta_i^k = \begin{cases} 1 \text{ if } C_i \cdot S^k - \theta \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

where θ is the threshold of the unit. The outputs of the logic units pass through variably weighted connections to serve as inputs to an output unit.

It is assumed that there are M input patterns, each to be grouped into one of two classes. With each pattern, there is an associated number δ which indicates the desired classification of the pattern. Thus

$$\delta^k = \begin{cases} 1 \text{ if } S^k \text{ is to be classified in the positive class} \\ -1 \text{ if } S^k \text{ is to be classified in the negative class} \end{cases}$$

A sequence of patterns is selected. The patterns are drawn from the population of patterns to be classified. Any individual pattern may appear in this sequence several times, or it need not appear at all. The number of appearances of the pattern S^j in the sequence is denoted by n^j. Following the selection of the connection vectors, this sequence of patterns is shown to the machine, during which the weights of the variable connections are modified.

The connections from the logic units to the report-out unit have variable weights. These weights are initially set to zero and then varied according to the following rule: if the i-th logic unit is activated when pattern S^k is shown during the adaptive sequence, then δ^k is added to the weight of the connection from the i-th unit to the report-out unit. Otherwise, the weight is unchanged. This can be restated as: the weight of the connection from the i-th unit to the report-out unit is changed by $\delta^k \delta_j^k$ each time S^k is shown during training. This is known as the "forced learning" rule.

Thus the weight of the i-th connection will be

$$w_i = \sum_{k=1}^{M} n^k \delta^k \delta_i^k$$

after the adaptive sequence. Consequently, if the input to the report-out unit when S^t is shown is denoted by β^t, then

$$\beta^t = \sum_{i=1}^{N} w_i \, \delta_i^t = \sum_{i=1}^{N} \sum_{k=1}^{M} n^k \delta^k \delta_i^k \delta_i^t$$

The output unit is said to have made a correct decision if

$$\beta^t \gtrless \theta_o$$

for

$$\delta^t = \pm 1 \text{ respectively.}$$

Consider the factors influencing β^t, and hence the classification of S^t:

δ^k = Previously determined classification for S^k

δ_i^k = Determined by connection vector of i-th logic unit

n^k = Choice of frequency of S^k in the adaptive sequence

Note that it follows that with a particular choice of network, adaptive sequence, and classification scheme, the decision on the test pattern, S^t, is completely deterministic.

When it is presumed that a random selection of connection vectors has been made, then β^t may be analyzed as a random variable.

Developing β^t as the sum of N independent random variables permits an evaluation of the variance:

$$\sigma^2(\beta^t) = \sum_{i=1}^{N} \sigma^2(\beta_i^t) = \sum_{i=1}^{N} \left\{ E(\beta_i^t)^2 - \left[E(\beta_i^t)\right]^2 \right\}$$

$$\beta_i^t = \sum_{k=1}^{M} n^k \delta^k \delta_i^k \delta_i^t$$

$$(\beta_i^t)^2 = \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k \delta_i^j \delta_i^k \delta_i^t$$

Defining the following symbols:

$Q_{jt} = E\left\{\delta_i^j \delta_i^t\right\} =$ Probability of selecting a connection vector such that the i-th associative unit is active for both S^j and S^t under the assumption of random selection of connection vectors, that Q_{jt} is independent of i.

$Q_{jkt} = E\left\{\delta_i^j \delta_i^k \delta_i^t\right\} =$ Probability of selecting a connection vector such that the i-th associative unit is active for S^j, S^k, and S^t, again independent of i.

One may write:

$$E\left(\frac{1}{N}\beta^t\right) = \sum_{j=1}^{M} n^j \delta^j Q_{jt} \qquad (1)$$

and

$$\sigma^2(\beta_i^t) = \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k \left\{Q_{jkt} - Q_{jt} Q_{kt}\right\}$$

Hence:

$$\sigma^2\left(\frac{1}{N}\beta^t\right) = \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k \left\{Q_{jkt} - Q_{jt} Q_{kt}\right\} \qquad (2)$$

Once a threshold for the response unit is selected, the mean and variance of β^t may be used in the Tschebycheff inequality to provide an absolute bound on the error probability. It may also be assumed that β^t is approximately normally distributed (since it is the sum of a large number of independent, well-behaved variables) to provide more reasonable estimates of the error probability.

One may also consider the test pattern S^t as having been selected at random either from the class of positive patterns (cloud formations containing vortices) or from the negative class. In this case, the variance of the input to the response unit should contain a component due to the selection of the test pattern. One obtains

$$E\left(\frac{1}{N}\beta_{\pm}\right) = \frac{1}{M_{\pm}} \sum_{t} \sum_{j=1}^{M} n^j \delta^j Q_{jt} \qquad (3)$$

and

$$\sigma^2\left(\frac{1}{N}\beta_{\pm}\right) = \frac{1}{NM_{\pm}} \sum_t \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k (Q_{jkt} - Q_{jt} Q_{kt})$$

$$+ \frac{1}{M_{\pm}} \sum_t \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k Q_{jt} Q_{kt}$$

$$- \left[\frac{1}{M_{\pm}} \sum_t \sum_k n^k \delta^k Q_{kt}\right]^2 \qquad (4)$$

where the summations on t are restricted to patterns in the appropriate class. The variance has two components: the first is the average of the variances of β^t and may be made arbitrarily small by increasing the number of logic units; the second is the variance of $E(\beta^t)$ over the selection of a test pattern, and is not affected by a change in the number of logic units. The second component limits the performance which may be achieved with a perceptron.

Approximate Perceptron Analysis

The computation of the exact variance of β^t is very difficult because the numbers Q_{jkt} are quite numerous and are not readily available. Q_{jkt} may be obtained by a process similar to the one described in the following section for Q_{jt}, but a vastly expanded measurement program would be required. An approximate method of estimating the variance of β^t is thus desired.

An analysis of the terms contributing to the variance in Equation 2 indicates that the most significant ones occur when j=k. The approximation

$$Q_{jkt} \cong \frac{Q_{jk} Q_{kt}}{Q_k} \qquad (5)$$

is exact when j=k. Substituting Equation 5 into 2,

$$\sigma^2\left(\frac{1}{N}\beta^t\right) \cong \frac{1}{N} \sum_j \sum_k n^j n^k \delta^j \delta^k \left[\frac{Q_{jk} Q_{kt}}{Q_k} - Q_{jt} Q_{kt}\right] \qquad (6)$$

When the test pattern is to be selected randomly,

$$\sigma^2\left(\frac{1}{N}\beta_{\pm}\right) = \frac{1}{NM_{\pm}}\sum_t \sum_{j=1}^{M}\sum_{k=1}^{M} n^j n^k \delta^j \delta^k \left(\frac{Q_{jk}}{Q_k} - Q_{jt}\right)$$

$$+ \frac{1}{M_{\pm}}\sum_t \sum_{j=1}^{M}\sum_{k=1}^{M} n^j n^k \delta^j \delta^k Q_{jt} Q_{kt}$$

$$- \left[\frac{1}{M_{\pm}}\sum_t \sum_k n^k \delta^k Q_{kt}\right]^2 \quad (7)$$

where the summations on t are restricted to patterns in the positive class or negative class, as appropriate. The expressions in Equations 3 and 7 may be rewritten as

$$E\left(\frac{1}{N}\beta_{+}\right) = \frac{1}{M_{+}}\sum_t \left[M_{+}\tilde{Q}_{t+} - M_{-}\tilde{Q}_{t-}\right] \quad (8a)$$

$$E\left(\frac{1}{N}\beta_{-}\right) = \frac{1}{M_{-}}\sum_t \left[M_{+}\tilde{Q}_{t+} - M_{-}\tilde{Q}_{t-}\right] \quad (8b)$$

and

$$\sigma^2\left(\frac{1}{N}\beta_{+}\right) = \left[\frac{1}{NM_{+}}\sum_{k=1}^{M} n^k \delta^k M_{+} Q_{k+}\left(\frac{M_{+}\tilde{Q}_{k+} - M_{-}\tilde{Q}_{k-}}{Q_k}\right)\right.$$

$$\left. - \sum_t \left(M_{+}\tilde{Q}_{t+} - M_{-}\tilde{Q}_{t-}\right)^2\right]$$

$$+ \frac{1}{M_{+}}\sum_t \left(M_{+}\tilde{Q}_{t+} - M_{-}\tilde{Q}_{t-}\right)^2$$

$$- \left[\frac{1}{M_{+}}\sum_t \left(M_{+}\tilde{Q}_{t+} - M_{-}\tilde{Q}_{t-}\right)\right]^2 \quad (9a)$$

$$\sigma^2\left(\frac{1}{N}\beta_{-}\right) = \frac{1}{NM_{-}}\left[\sum_{k=1}^{M} n^k \delta^k M_{-}\tilde{Q}_{k-}\left(\frac{M_{+}\tilde{Q}_{k+} - M_{-}\tilde{Q}_{k-}}{Q_k}\right)\right.$$

$$\left. - \sum_t \left(M_{+}\tilde{Q}_{t+} - M_{-}\tilde{Q}_{t-}\right)^2\right]$$

$$+ \frac{1}{M_{-}}\sum_t \left(M_{+}\tilde{Q}_{t+} - M_{-}\tilde{Q}_{t-}\right)^2$$

$$- \left[\frac{1}{M_{-}}\sum_t \left(M_{+}\tilde{Q}_{t+} - M_{-}\tilde{Q}_{t-}\right)\right]^2 \quad (9b)$$

where the summation on t is again restricted to patterns of the appropriate class. The symbols $\tilde{Q}_{k+}$ and $\tilde{Q}_{k-}$ are used to indicate the average values of Q_{kt} for t in the positive class and negative class respectively.

Estimation Technique

This section describes how the approximations of Equations 8 and 9 of the preceding section may be combined to estimate the number of logic units required to achieve a given performance level, and the means by which the required parameters may be estimated using optical correlations of sample patterns.

The input to the report-out unit is a random variable due to the random nature of the selection of the logic units, and the random selection of the test pattern. The conditional mean and variance of the input given that the test pattern is a vortex pattern, and the conditional mean and variance given that the pattern does not contain a vortex, may be computed using Equations 8 and 9. For a given test pattern, the input to the output unit is approximately normally distributed, since it is the sum of a large number of independent variables — the outputs of the logic units. If the number of patterns is large, and if the means within a class are nearly normally distributed, this approximation may be extended to a randomly selected test pattern. Given a threshold for the response unit, the conditional means and variances may be used to compute the false alarm (classifying a pattern as a vortex when it is not) and missed target rates (not classifying a pattern as a vortex when it is).

The evaluation of Equations 8 and 9 is dependent on the evaluation of the quantities $\tilde{Q}_{t+}$ and $\tilde{Q}_{t-}$ for various values of t. This may be accomplished by evaluating Q_{jt} for all values of the subscript j, or by sampling the possible range of this index, as was done in this study. A method is thus required for estimating Q_{jt}, the probability that a logic unit will be activated by pattern S^j and by pattern S^t.

Estimation of Q_{jt} Using Optical Correlation

Let F_{jt} (u, v) denote the bivariate distribution function of the transparencies S^j and S^t. The contribution of a given input connection to the total input to the logic unit may be regarded as being equal in magnitude to the transparency of the pattern at the origin point of the connection, times the weight of that connection. If the origins of the connections ar chosen at random, independently and according to a uniform distribution, the following analysis applies.

Denote by X_{ji} the input to the i-th connection when S_j is shown. X_{ji} and X_{ti} have the joint distribution function $F_{jt}(X_{ji}, X_{ti})$ - the same distribution function as the transparencies. The joint distribution of the total inputs X_j and X_t is given by the convolution of connection density functions, since the origin points are selected independently. Denote this distribution by $G_{jt}(X_j, X_t)$.

Distinguishing those parameters pertaining to the distribution F_{jt} by tildas (e.g., $\tilde{m}_j$) we have

$$m_j = A\,\tilde{m}_j \qquad \sigma_{jt} = B\,\tilde{\sigma}_{jt}$$

$$m_t = A\,\tilde{m}_t \qquad \sigma_t^2 = B\,\tilde{\sigma}_t^2$$

$$\sigma_j^2 = B\,\tilde{\sigma}_j^2$$

Thus it is necessary to estimate $\tilde{m}_j$, $\tilde{m}_t$, $\tilde{\sigma}_t^2$, and $\tilde{\sigma}_{jt}$ using optical meansurements.

$$\tilde{m}_j = \int_{u,v} u\, d\, F_{jt}(u,v) = \int_u u\, d\, F_j(u)$$

Hence $\tilde{m}_j$ is given by the total light passed by the S^j transparency, normalized by dividing by the total light passed by an empty frame. $\tilde{m}_t$ is obtained similarly.

$$\tilde{\sigma}_j^2 = \int_{u,v} u^2\, d\, F_{jt}(u,v) - \tilde{m}_j^2$$

$\tilde{\sigma}_j^2$ is obtained by measuring the total light passed by two S^j transparencies exactly superimposed, normalizing as above, and subtracting $\tilde{m}_j^2$.

$$\tilde{\sigma}_{jt} = \int_{u,v} u\, v\, d\, F_{jt}(u,v) - \tilde{m}_j\, \tilde{m}_t$$

$\tilde{\sigma}_{jt}$ is obtained by subtracting $\tilde{m}_j \tilde{m}_t$ from the normalized quantity of light passed by superimposing the S^j and S^t transparencies. The correlation coefficient ρ is given by $\tilde{\sigma}_{jt}/\tilde{\sigma}_j \tilde{\sigma}_t$ for both F_{jt} and G_{jt}.

The calculation of Q_{jt} is greatly simplified by the assumption that the total input to a logic unit when a pattern is shown is approximately normally distributed. The pair (X_j, X_t) denoting the total input for patterns S_j and S_t, respectively, then would have a bivariate normal distribution.

The derivation of m_j, m_t, σ_t^2, and ρ is in no way dependent on the normality assumptions; the asumption only insures the sufficiency of this set of parameters.

The normality assumption is, then, that

$$dG_{jt}(X_j, X_t) \sim \frac{1}{2\pi\sigma_j\sigma_t\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{1}{1-\rho^2}\left[\frac{(X_j-m_j)^2}{(\sigma_j)^2} - 2\rho\frac{(X_j-m_j)}{(\sigma_j)}\frac{(X_t-m_t)}{(\sigma_t)} + \frac{(X_t-m_t)^2}{(\sigma_t)^2}\right]} dX_j dX_t$$

Given the normal approximation, and the requisite parameters, the next problem is to obtain Q_{jt}.

$$Q_{jt} = \frac{1}{2\pi\sigma_j\sigma_t\sqrt{1-\rho^2}} \int_\theta^\infty \int_\theta^\infty e^{-\frac{1}{2}\frac{1}{1-\rho^2}\left[\frac{(X_j-m_j)^2}{(\sigma_j)^2} - 2\rho\frac{(X_j-m_j)}{(\sigma_j)}\frac{(X_t-m_t)}{(\sigma_t)} + \frac{(X_t-m_t)^2}{(\sigma_t)^2}\right]} dX_j dX_t$$

By substitution, and letting $a = \frac{\theta - m_j}{\sigma_j}$ and $b = \frac{\theta - m_t}{\sigma_t}$

$$Q_{jt} = \int_b^\infty \int_a^\infty \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{1}{1-\rho^2}\left[y^2 - 2\rho yz + z^2\right]} dy dz \qquad (10)$$

Many forms for integrating (10) numerically exist[1]. The derivation of the form used in this study is documented in the final report[2]. The results are summarized as follows for various ranges of a and b.

---

[1] Gupta, Shanti S. (1963). "Probability Integrals of Multivariate Normal and Multivariate t," Ann. Math. Stat., 34, 792-838.

[2] "Cloud Pattern Interpretation," Astropower, Inc. Report 129-F, August 1963.

a	b	Q_{jt}
0	0	$\frac{1}{\pi} \tan^{-1} \sqrt{\frac{1+\rho}{1-\rho}}$
> 0	0	$F(0, a, \rho)$
< 0	0	$\frac{1}{2} + F(0, a, \rho)$
0	> 0	$F(0, b, \rho)$
0	< 0	$\frac{1}{2} + F(0, b, \rho)$
< 0	< 0	$1 + F(a, b, \rho) + F(b, a, \rho)$
All Others		$F(a, b, \rho) + F(b, a, \rho)$

where

$$F(\xi, \eta, \rho) = \frac{\eta e^{-\frac{\eta^2}{2}}}{2\pi} \int_{\frac{\xi - \eta\rho}{\sqrt{1-\rho^2}}}^{\infty} \frac{e^{-\frac{u^2}{2}}}{u^2 + \eta^2}\, du$$

OPTICAL CORRELATION MEASUREMENTS

More than 1200 wide angle photos from TIROS V and TIROS VI were obtained directly from the TIROS Data Acquisition Station on Wallops Island. All photos that appeared to show the familiar vortex pattern were selected. For each photo showing a vortex, a second was selected showing no obvious storm formation.

A further selection was made considering the area of earth coverage within the photo. Since it is conceivable that the selected vortex photos could show, in general, more horizon in a certain picture area than would nonvortex photos, or vice-versa leading to an incorrect interpretation of the classification clues, it was decided to exclude horizon areas by selecting only a round disc within each rectangular negative. This resulted in the additional benefit that any two pictures to be correlated could be rotated with respect to each other without changing the overlap area. Thus more than one measurement could be obtained for each pair with little added effort.

The selected 35mm cloud cover negatives had to be copied onto circular slides that could be inserted in a pair of rotatable slide holders in the optical correlator. To obtain the necessary mechanical rigidity, the slide frames were fabricated from hard aluminum. An inner diameter of 1 in. was chosen for the transparency area.

An identical pair of slides for each photographic negative was required to find its mean gray level and its gray level variance. To obtain identical pairs, unexposed photographic material was bonded to the frames. Then a number of frames were exposed in sequence to the enlarged image of the original 35mm negative. Exposure time, brightness, development time, and developer liquid were held as nearly constant as possible. In this way for each negative two frames were obtained that were identical in their total transmittance within 5%. The final result of this photographic processing was 50 slide pairs of the vortex class and 50 slide pairs of the nonvortex class. There now existed a combination of 5000 picture pairs that could be correlated against each other. Since the rotation of a slide in the image plane results in another pattern with respect to the envisioned recognition system, the number of possible pattern combinations could be increased indefinitely. For a realistic compromise it was decided that a total of approximately 10,000 measurements would provide a reasonable estimate of the overlap areas between the pattern classes as required for the mathematical analysis.

The following measurements were conducted:

a. 100 measurements of the transmittance of an exactly aligned pair. After these measurements were performed, only one slide of each pair was used for further measurements.

b. It was decided that each negative should be correlated to 16 other negatives of its own class and 16 of the opposite class, thus resulting in 1600 pairs. Each pair was to be measured for six different rotations relative to each other, yielding 9600 measurements.

A logbook was designed which enumerated the individual picture pairs to be correlated. The vortex negatives were numbered 1 to 50, the nonvortex negatives 51 to 100. A 100 x 100 square matrix was then plotted and individual squares were marked. A mark at the crossing of column m with row n indicates that negatives m and n form one of the picture pairs to be correlated. The design was started with the upper left quarter (established by the overlap of the first 50 rows with the first 50 columns), which signifies vortex-vortex correlations. After excluding the diagonal from 1-1 to 50-50, marks were inserted at random, starting at the upper left corner, with the restrictions that each column and each row should contain 16 marks, and that the resulting pattern should be symmetrical about the diagonal. The same procedure was applied to the lower two quarters using the pattern already established in the first quarter and reshuffling numbers at random. The resulting design assured a good spread of picture combinations. From the chart a log book was compiled which, after insertion of the measurement data was transferred directly to IBM cards by a keypunch operator.

A system schematic of the optical correlator is shown in Figure 5. Figure 6 gives an overall view of the equipment as used in making the measurements. A modulated light beam is generated in the small black enclosure mounted on the left end of the optical bench. The diverging beam emerges from an aperture in the front of the source enclosure and passes through a collimating lens.

TP1967

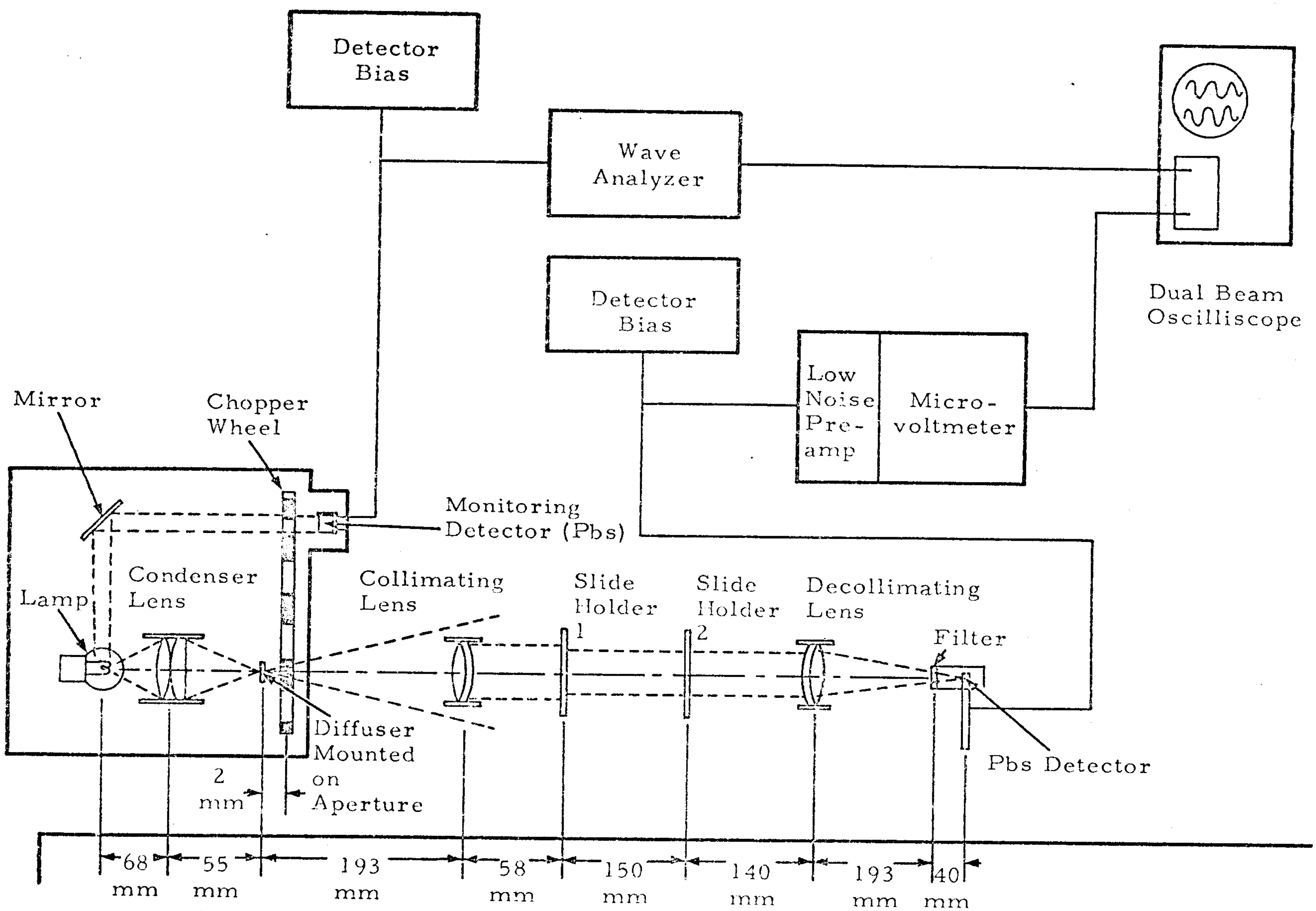


Figure 5. System Schematic

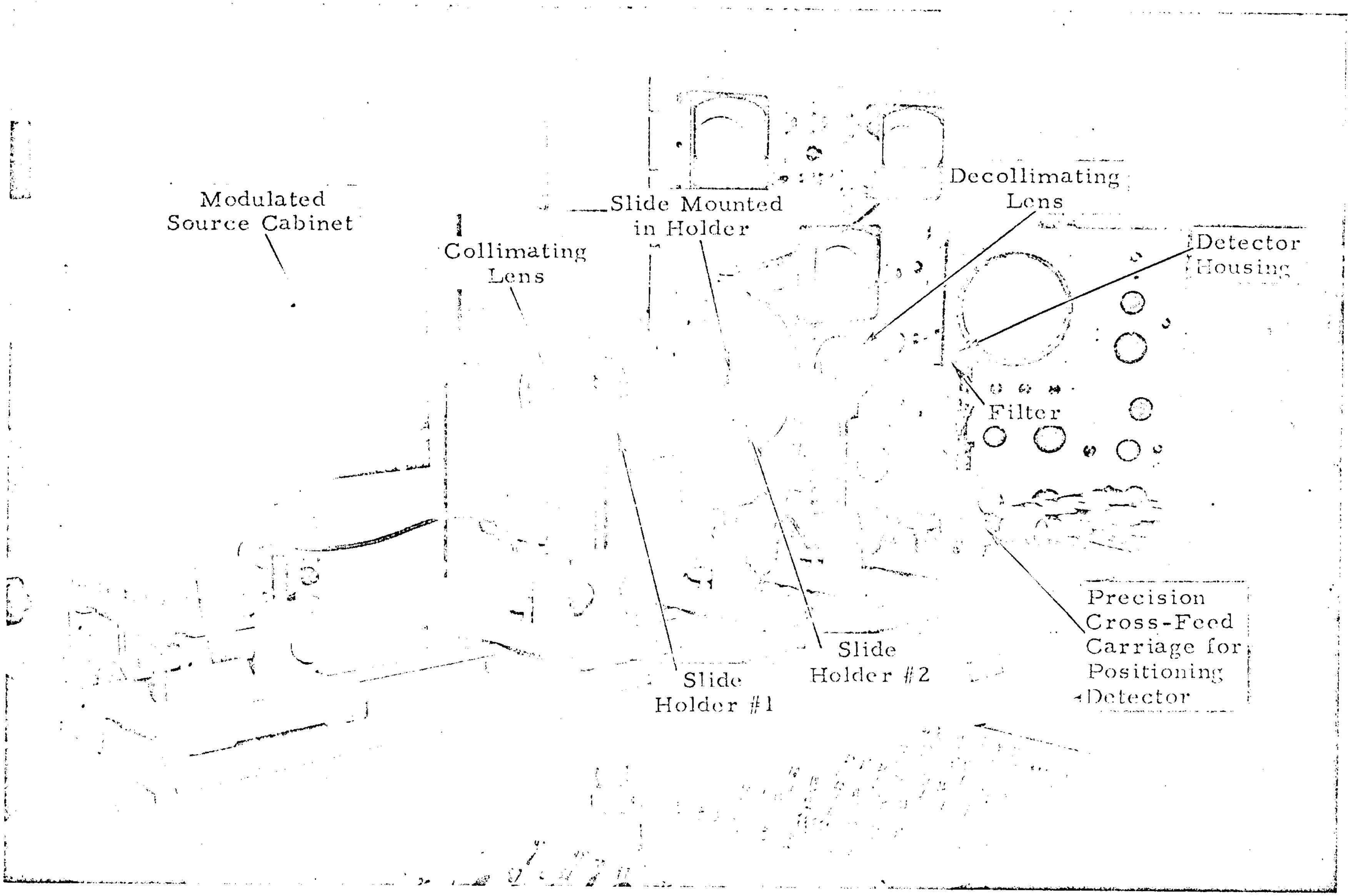


Figure 6. Arrangement of Components on Optical Bench

The collimated beam then passes through the transparencies mounted in the two slide holders. After passing through the transparencies, the collimated beam is decollimated, filtered and focused upon a lead sulfide detector. The output of the photo detector is measured by a tunable microvoltmeter. The reading of the microvoltmeter is directly proportional to the combined transmittance of the slide pair inserted into the holders.

The setup also contained a monitoring channel for monitoring constant intensity of the light source. The stability of the correlator was further checked by periodic calibration measurements to insure constant chopper frequency (240 cps), exact optical alignment, constant energy density over the slide area ($\pm 2\%$ tolerance was obtained), and linearity of the overall system over the full range of output amplitudes.

For the correlation measurement of each pair one slide was placed in slide holder #2 and the other slide was placed in slide holder #1, which was capable of 360° rotation in the plane of the slide. The slide holder was engraved with degree graduation so that the slide could be accurately positioned within $1/2^\circ$ of any designated rotational position.

Six measurements were made for each pair of slides inserted into the holders. For all six measurements, the slide in holder #2 (closest to the detector) was inserted and fixed with its index mark at 0°. For the first measurement, the slide in holder #1 was positioned so that its index mark was at 0°. The slide was then rotated 60° for each of the remaining five measurements. Upon completion of the six measurements, the pair of slides was replaced by the next pair, as designated by the logbook. This procedure was followed until the 1600 sets of measurements were completed.

THE COMPUTER INVESTIGATION

System Parameters Under Investigation

The system under investigation is specified to be a forced learning perceptron. The logic units in the perceptron are specified to have their input connections arising from sensory points selected at random according to a uniform distribution. A representative set of patterns was selected. The remaining system parameters are:

a. Logic unit parameters
b. Number of logic units
c. Output unit threshold

The computer investigation may be characterized as a search for a suitable set of logic unit parameters. The optimum threshold for the output unit and the number of logic units required to approach asymptotic performance were obtained for the various logic unit parameter sets.

The logic unit parameters of interest are the number of input connections per logic unit, the weights assigned to the connections, and the logic unit threshold. The logic unit parameters are completely described by the two numbers:

$$\alpha = \frac{\Sigma w_i}{\sqrt{\Sigma w_i^2}}$$

$$\beta = \frac{\theta}{\sqrt{\Sigma w_i^2}}$$

where the w_i's are the weights assigned to the connections (and may be positive or negative) and θ is the unit threshold.

General Form of Program and Sequence of Computer Runs

To perform the estimation procedure, the equations of the approximate perceptron analysis were mechanized on an IBM 7094. The computer routine consists of a main program which examined the data obtained by the optical correlation procedure, and two subroutines. One subroutine computed the empirical distribution of Q_{jt} from the equations with limits imposed by the patterns being tested and the other provides a means for printing out some of the results in a suitable format. A simplified flow chart of the computer routine is shown in Figure 7.

The 9600 data points were processed for a number of parameter combinations. The purpose of these repeated trials was to attempt to find an optimum for the input connections to the logic units and a logic unit threshold, one that would process the data with a minimum of errors. Each computer run consisted of seven parameter combinations and took approximately 10 minutes of computer time.

In order to permit conducting a more extensive investigation, the possibility of using a fraction of the available data was investigated. The program checkout decks involving 120, 720, and 840 optical correlation measurements were completely inadequate. A set of 2400 measurements was carefully selected to provide 24 measurements on Q_{k+} and 24 measurements on Q_{k-} for each value of the index k. The results obtained with these short data decks were compared with those obtained using the full set of 9600 correlations (96 observations on each Q_{k+} and Q_{k-}). The means and standard deviations returned by the short deck generally differed by 20 to 50% from the values given by the full deck. This accuracy was considered inadequate, and the use of partial data decks was discontinued.

Description of Program

The implementation of the program consisted of the following:

a. Read in half of C table (integral of the normal distribution), compute other half of C table. Compute μ^2 and $\dfrac{e^{\frac{-\mu^2}{2}}}{\sqrt{2\pi}}$.

b. Compute $D_1 = n_x - n_y$ and $D_2 = \sqrt{n_x + n_y}$, multiplying factors for the mean and standard deviation of the patterns of one input to convert to many inputs.

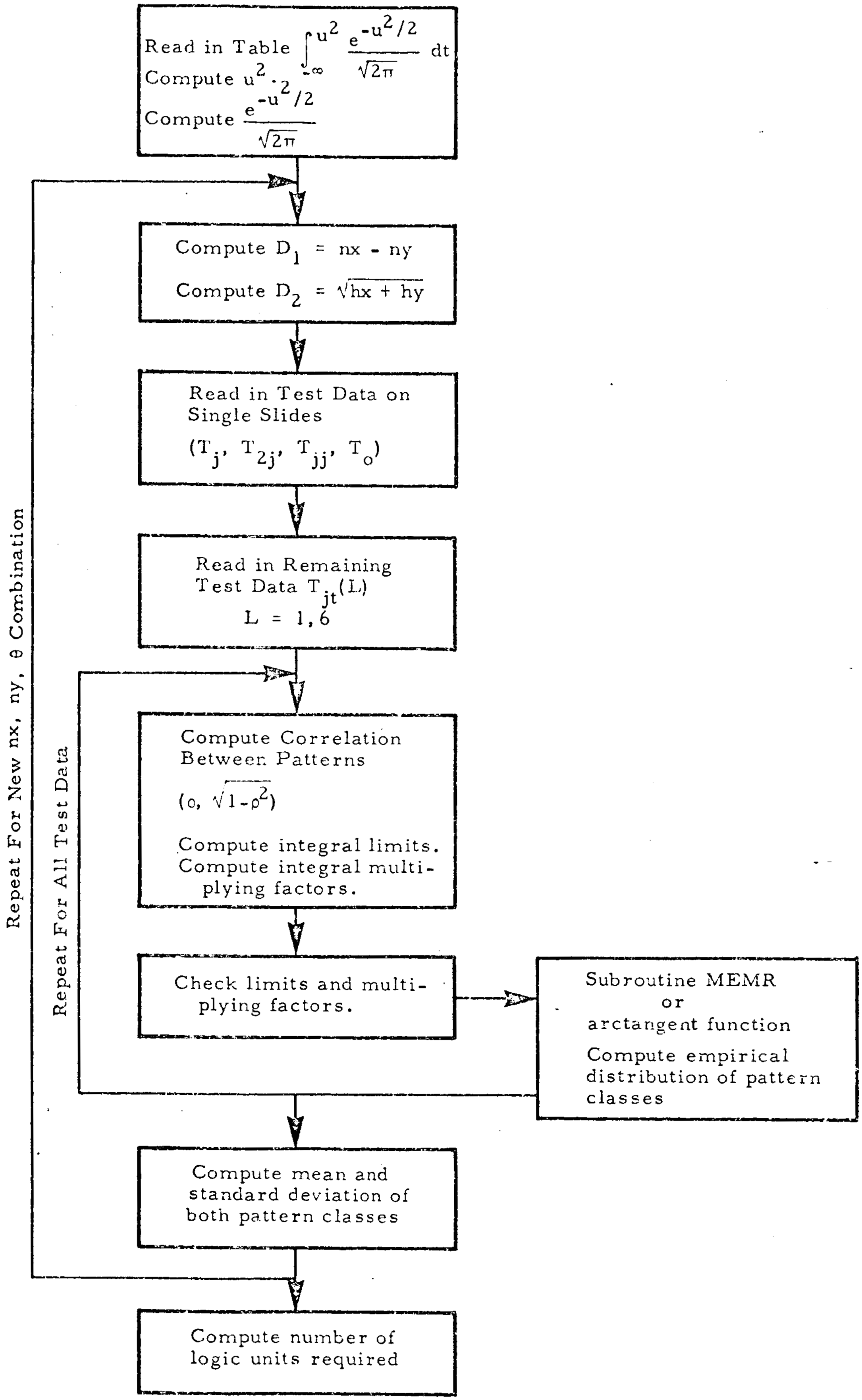


Figure 7. Simplified Flow Chart of Computer Program

c. Read in the single slide (T_{1j} and T_{2j}) and the single slide squared (T_{jj}) measurements and normalized by ($1/T_o$).

d. Read in remaining optical correlation measurements (T_{jt}), six rotations per slide pair, normalized by ($1/T_o$).

NOTE: Since storage restraints did not permit operating on all data points during one pass through the computer, the correlations were performed in groups of 20 data cards per pass.

e. Determine the class of the pattern under consideration (from its index number).

f. Compute the correlation (ρ) between the two patterns being processed for each rotation and $w = \sqrt{1-\rho^2}$.

g. Compute the limits of the integral.

h. Compute Q_j and Q_t (single-slide probability of activating a logic unit).

i. Check the limits to determine if they are above or below the bounds set on the integral, or equal to zero. Check integral multiplying factors $\neq 0$.

j. Compute the empirical distribution of Q_{jt}.

k. Compute the average of Q_{jt} (Q_{k+} and Q_{k-}) for patterns in each class.

l. Compute the mean and the standard deviation for both pattern classes. Compute the standard deviation as a function of both the selection of the logic units and the distribution of the patterns.

m. Compute the difference of the means.

n. Compute the separation of pattern classes as a function of the number of logic units in the machine.

o. Select threshold which minimizes the number of errors on the asymptotic performance.

The computer routine returned the following outputs:

a. The expected values of the input to the decision element for patterns of the positive class and for patterns of the negative class, and the difference between these values.

b. The inherent standard deviation within each class.

c. The standard deviation for each class due to logic unit selection.

d. The asymptotic class separation in standard deviations (see Figure 8).

e. The minimum number of misclassified patterns, asymptotic performance (see Figure 7).

f. A table of the expected input to the decision element as a function of the pattern.

g. A table of the class separation in standard deviations, as a function of the number of logic units (see Figure 9).

h. The decision unit threshold producing minimum errors in asymptotic performance, and a table of these errors.

i. A table of the probability of a logic unit being active, as a function of the pattern.

j. Tables of Q_{k+} and Q_{k-}.

k. Tables of the empirical distribution of Q_{jk} for two vortex patterns, one vortex and one nonvortex pattern, and two non-vortex patterns.

Results of Computer Program

Using the final program and the full data set, some 77 combinations of α and β were investigated. In addition, 36 other combinations were investigated (16 with the full data deck) with an earlier version of the program. The earlier program did not provide all of the outputs desired, but was adequate to determine the suitability of the parameters.

TP1967

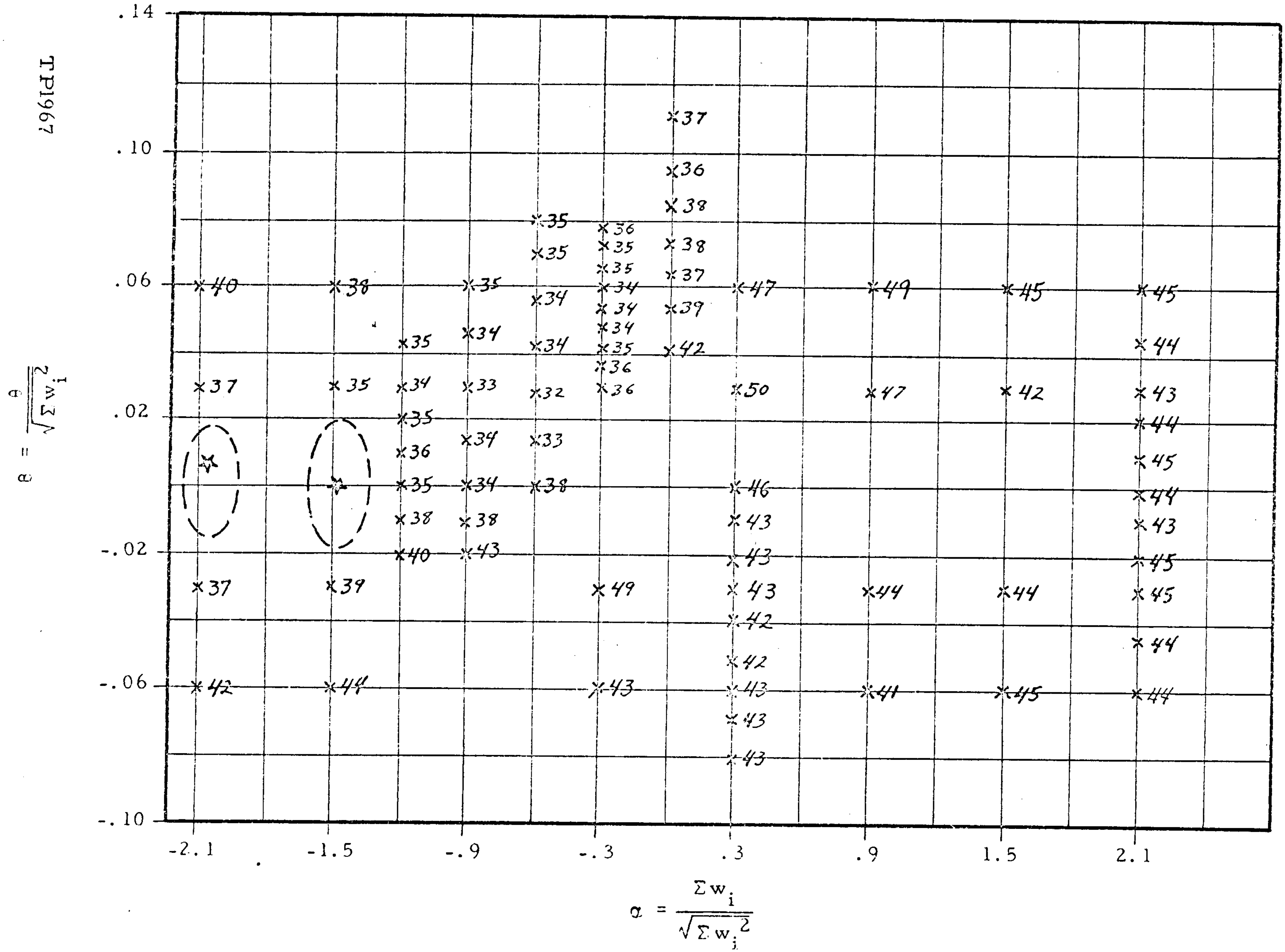


Figure 8. Minimum Number of Errors

TP1967.

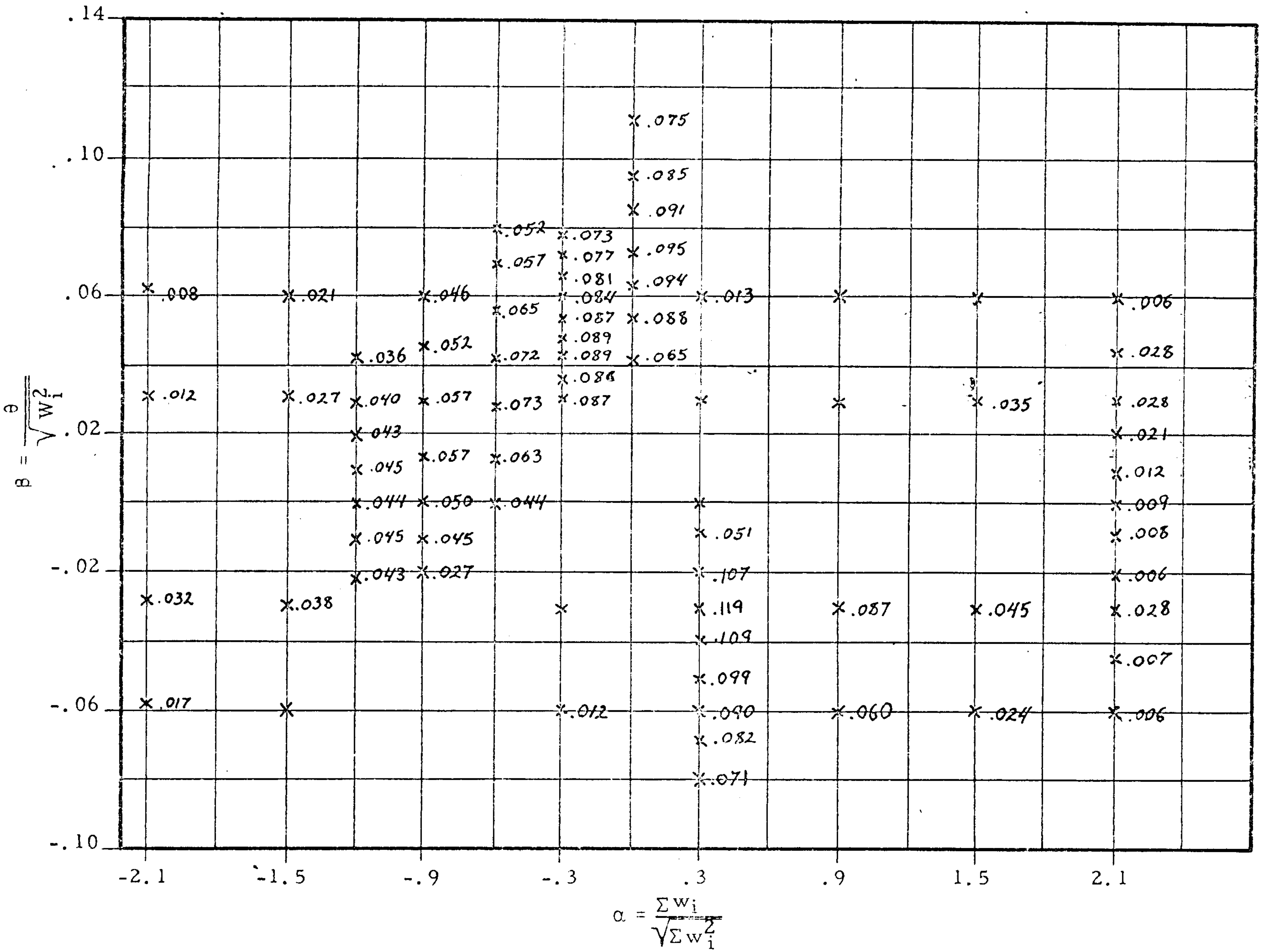


Figure 9. Asymptotic Class Separation in Standard Deviations

The 77 points were obtained in three passes on the computer. In the first pass, 28 scattered points were selected, the choice being guided by the earlier partial results on 36 points. Based on these 28 combinations, 28 more points were selected to provide a finer grid in certain areas. The third pass of 21 points examined the region around $\alpha = 0.9$ and $\beta = 0.0$, where the number of errors appeared to be approaching a minimum.

None of the parameter combinations resulted in good separation of the pattern classes. The approximation used in estimating the component of variance due to logic unit selection requires good class separation for high accuracy. In many cases, separation was so poor that negative estimates of this component of variances were returned. In some cases, the mean value of the input to the response unit for the positive class was less than that for the negative class. Two criteria were used in deciding whether a parameter combination warranted further consideration:

a. The mean value for the positive class had to be greater than the mean value for the negative class.

b. The variance estimates had to be positive.

Only 11 of the 28 combinations in the first pass met these requirements. (It can be shown that failure to meet condition (a) is an inadequacy of the optical correlation, and not of the perceptron technique. If good separation had been obtained, however, the measurement inaccuracies would not be of such significance.)

The most significant computer outputs are summarized in Figures 8 through 10. For each parameter set a threshold for the output unit was selected to produce a minimum number of errors in the asymptotic performance. These minimum values are shown in Figure 8. The lowest value obtained was 32. A second threshold for the output unit was selected, this one providing maximum number of standard deviations for the threshold-to-class mean separation. These separations again are for the asymptotic performance. They are presented in Figure 9. The absence of an entry on the chart indicates failure to meet requirement (a) above. The number of logic units required to achieve separations equal to 99.9% of the asymptotic separation are given in Figure 10. Since the computer must return positive estimates of the variance for these calculations, Figure 10 reveals those parameter combinations for which the approximation failed to yield positive variances.

Summarizing the results, parameter combinations which could be analyzed at all were difficult to find. For those combinations which could be analyzed, class spearations were poor, giving rise to high error rates. A network of about 400 to 500 logic units should provide essentially asymptotic performance for the better parameter sets. However, if the interclass separations were better (a factor of 20 seems desirable) the error rate could be reduced significantly. More logic units would then be required to achieve asymptotic performance.

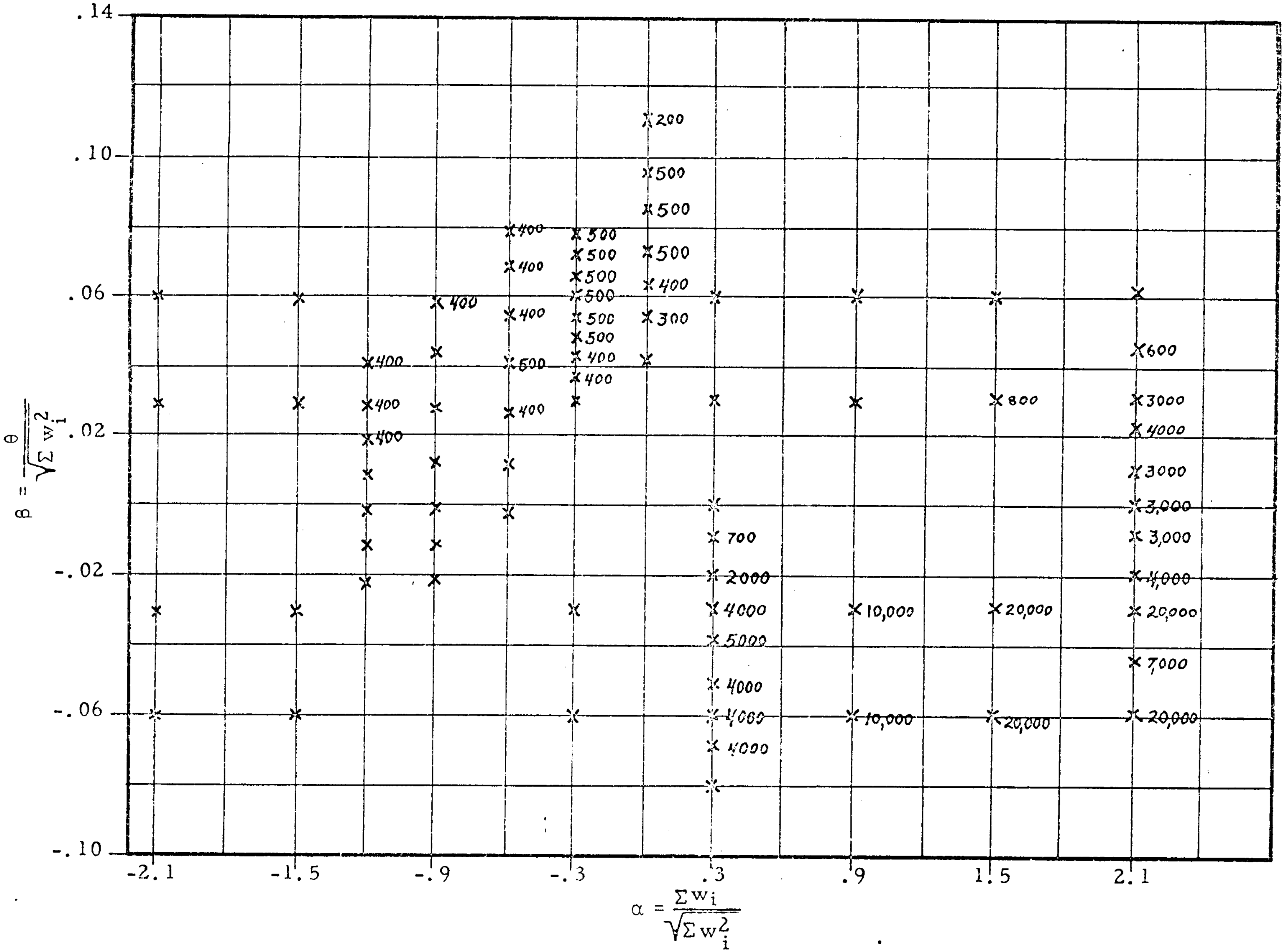


Figure 10. Number of Logic Units Required for 99.9% of Limiting Separation

CONCLUSIONS

The primary purpose of this program was to show the feasibility of the application of self-organizing techniques to the design of a parallel logic system for the automatic interpretation of satellite cloud cover photographs. This was to be accomplished in two stages:

a. The estimation of machine size required by the application of a learning algorithm that had been thoroughly documented mathematically - the forced learning procedure. The estimate was to be based on the results of optical correlation of a number of actual TIROS photographs containing both storm and nonstorm components. Only gross storm features (vortex structures) were to be involved in this first estimation task.

b. The extrapolation of the results obtained from the application of this open-loop, nonselective, random property algorithm to more powerful algorithms, such as natural selection and non-random generation of the property list, to increase logic unit efficiency and reduce machine size.

Since it was believed that a self-organizing system using forced learning could be made to asymptotically approach perfect performance for a recognition task of this nature, economic feasibility then could be determined simply by estimating the required complexity of the logic layer (in terms of the number of logic units) to arrive at some preselected closeness of approach to perfect performance.

Program Results

Optical correlations on 100 cloud cover transparencies (50 vortex, 50 nonvortex) were performed with a number of rotations of each pattern, giving a total of 9600 data points. This data was examined by a computer routine which mechanized the estimation procedure. Results of this program may be summarized as follows:

a. The nature of the patterns to be classified was such that good separation between pattern classes (vortex and nonvortex) was not obtainable. This was due in part to the inaccuracy of the optical correlation procedure and photographic processing (5% overall accuracy was achieved through diligent photograph selection and control of the measurement apparatus). However, the primary cause appeared to be the complexity of the patterns themselves.

b. Computer analysis of the resulting test data showed that due to the lack of good class separation, the limiting performance of a forced learning perceptron was in the order of 65 to 70% of perfect performance. This level of performance could be achieved with 400 to 500 logic units.

c. The study of alternative self-organizing routines has produced more powerful algorithms involving a closed-loop decision process which promises considerable improvement of system performance by incorporating a learning operation having essentially two stages. The first stage is the implementation of a nonrandom procedure for the generation of the property list (discriminant analysis). The second stage is the maximization of the class separation by a closed-loop learning process which concentrates on those patterns most difficult to classify (iterative design).

ACKNOWLEDGMENTS

The authors wish to thank Mr. H. Lowell of NASA Headquarters for his encouragement and valuable suggestions during this program and Mr. S. Spinak of the Douglas Optical Laboratory in Santa Monica who achieved the optical correlation measurements.